

高中数学选修 2-3 知识点

第一章 计数原理

1.1 分类加法计数与分步乘法计数

分类加法计数原理：完成一件事有两类不同方案，在第 1 类方案中有 m 种不同的方法，在第 2 类方案中有 n 种不同的方法，那么完成这件事共有 $N=m+n$ 种不同的方法。分类要做到“不重不漏”。

分步乘法计数原理：完成一件事需要两个步骤。做第 1 步有 m 种不同的方法，做第 2 步有 n 种不同的方法，那么完成这件事共有 $N=m \times n$ 种不同的方法。分步要做到“步骤完整”。

n 元集合 $A=\{a_1, a_2, \dots, a_n\}$ 的不同子集有 2^n 个。

1.2 排列与组合

1.2.1 排列

一般地，从 n 个不同元素中取出 $m(m \leq n)$ 个元素，按照一定的顺序排成一列，叫做从 n 个不同元素中取出 m 个元素的一个排列(**arrangement**)。

从 n 个不同元素中取出 $m(m \leq n)$ 个元素的所有不同排列的个数叫做从 n 个不同元素中取出 m 个元素的排列数，用符号 A_n^m 表示。

排列数公式：

$$A_n^m = \frac{n!}{(n-m)!} = n(n-1)(n-2) \cdots (n-m+1)$$

n 个元素的全排列数

$$A_n^n = n!$$

规定： $0!=1$

1.2.2 组合

一般地，从 n 个不同元素中取出 $m(m \leq n)$ 个元素合成一组，叫做从 n 个不同元素中取出 m 个元素的一个组合(**combination**)。

从 n 个不同元素中取出 $m(m \leq n)$ 个元素的所有不同组合的个数，叫做从 n 个不同元素中取出 m 个元素的组合数，用符号 C_n^m 或 $\binom{n}{m}$ 表示。

组合数公式：

$$\therefore A_n^m = C_n^m \cdot A_m^m$$

∴

$$C_n^m = \frac{A_n^m}{A_m^m} = \frac{n!}{m!(n-m)!} = \frac{n(n-1)(n-2)\cdots(n-m+1)}{m!}$$

规定: $C_n^0 =$

组合数的性质:

$$C_n^m = C_n^{n-m} \quad (\text{“构建组合意义”——“殊途同归”})$$

$$C_{n+1}^m = C_n^m + C_n^{m-1} \quad (\text{杨辉三角})$$

$$kC_n^k = nC_{n-1}^{k-1}$$

$$*C_n^k \times C_{n-k}^{m-k} = C_n^m \times C_m^k$$

1.3 二项式定理

1.3.1 二项式定理(binomial theorem)

$$(a+b)^n = C_n^0 a^n + C_n^1 a^{n-1} b + \cdots + C_n^k a^{n-k} b^k + \cdots + C_n^n b^n \quad (n \in N^*)$$

其中各项的系数 C_n^k ($k \in \{0, 1, 2, \dots, n\}$) 叫做二项式系数(binomial coefficient);

式中的 $C_n^k a^{n-k} b^k$ 叫做二项展开式的通项, 用 T_{k+1} 表示通项展开式的第 $k+1$ 项:

$$T_{k+1} = C_n^k a^{n-k} b^k$$

*注意二项展开式某一项的系数与这一项的二项式系数是两个不同的概念。

1.3.2 “杨辉三角”与二项式系数的性质

*表现形式的变化有时能帮助我们发现某些规律!

(1) 对称性

(2) 当 n 是偶数时, 共有奇数项, 中间的一项 $C_n^{\frac{n}{2}+1}$ 取得最大值;

当 n 是奇数时, 共有偶数项, 中间的两项 $C_n^{\frac{n-1}{2}}$, $C_n^{\frac{n+1}{2}}$ 同时取得最大值。

(3) 各二项式系数的和为

$$2^n = C_n^0 + C_n^1 + C_n^2 + \cdots + C_n^k + \cdots + C_n^n$$

(4) 二项式展开式中，奇数项二项式系数之和等于偶数项二项式系数之和：

$$C_n^0 + C_n^2 + C_n^4 + \dots = C_n^1 + C_n^3 + C_n^5 + \dots$$

(5) 一般地，

$$C_r^r + C_{r+1}^r + C_{r+2}^r + \dots + C_{n-1}^r = C_n^{r+1} \quad (n > r)$$

第二章 随机变量及其分布

2.1 离散型随机变量及其分布

2.1.1 离散型随机变量

随着试验结果变化而变化的变量称为**随机变量(random variable)**。

随机变量和函数都是一种映射，随机变量把随机试验的结果映为实数，函数把实数映为实数。试验结果的范围相当于函数的定义域，随机变量的取值范围相当于函数的值域。

所有取值可以一一列出的随机变量，称为**离散型随机变量(discrete random variable)**。

概率分布列(probability distribution series)，简称为**分布列(distribution series)**。

X	x_1	x_2	...	x_i	...	x_n
P	p_1	p_2	...	p_i	...	p_n

也可用等式表示：

$$P(X = x_i) = p_i, \quad i = 1, 2, \dots, n$$

根据概率的性质，离散型随机变量的分布列具有如下性质：

(1) $p_i \geq 0, \quad i=1, 2, \dots, n;$

(2) $\sum_{i=1}^n p_i = 1$

随机变量 X 的**均值(mean)**或**数学期望(mathematical expectation)**：

$$E(X) = x_1 p_1 + x_2 p_2 + \dots + x_i p_i + \dots + x_n p_n$$

它反映了离散型随机变量取值的平均水平。

随机变量 X 的**方差(variance)**刻画了随机变量 X 与其均值 E(X)的平均偏离程度

$$D(X) = \sum_{i=1}^n (x_i - E(X))^2 p_i$$

其算术平方根 $\sqrt{D(X)}$ 为随机变量 X 的 **标准差(standard deviation)**。

$$E(aX + b) = aE(X) + b$$

$$D(aX + b) = a^2 D(X)$$

若随机变量 X 的分布具有下表的形式，则称 X 服从 **两点分布(two-point distribution)**，并称 $p=P(X=1)$ 为成功概率。(两点分布又称 **0-1 分布**。由于只有两个可能结果的随机试验叫伯努利试验，所以两点分布又叫伯努利分布)

X	0	1
P	$1-p$	p

若 X 服从两点分布，则 $E(X) = p$ ， $D(X) = p(1 - p)$

一般地，在含有 M 件次品的 N 件产品中，任取 n 件，其中恰有 X 件次品，则

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad k=0, 1, 2, \dots, m$$

X	0	1	...	m
P	$\frac{C_M^0 C_{N-M}^{n-0}}{C_N^n}$	$\frac{C_M^1 C_{N-M}^{n-1}}{C_N^n}$...	$\frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$

其中 $m=\min\{M, n\}$ ，且 $n \leq N$, $M \leq N$, $n, M, N \in \mathbb{N}^*$

如果随机变量 X 的分布列具有上表的形式，则称随机变量 X 服从 **超几何分布(hypergeometric distribution)**。

2.2 二项分布及其应用

2.2.1 条件概率

一般地，设 A, B 为两个事件，且 $P(A)>0$ ，称

$$P(B|A) = \frac{P(AB)}{P(A)}$$

为在事件 A 发生的条件下，事件 B 发生的 **条件概率(conditional probability)**。
如果 B 和 C 是两个互斥事件，则

$$P(B \cup C|A) = P(B|A) + P(C|A)$$

2.2.2 事件的相互独立性

设 A, B 为两个事件，若

$$P(AB) = P(A)P(B)$$

则称事件 A 与事件 B 相互独立(mutually independent)。

可以证明，如果事件 A 与 B 相互独立，那么 A 与 \bar{B} , \bar{A} 与 B , \bar{A} 与 \bar{B} 也都相互独立。

2.2.3 独立重复试验与二项分布

一般地，在相同条件下重复做的 n 次试验称为 n 次独立重复试验(independent and repeated trials)。

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

其中 A_i ($i=1, 2, \dots, n$) 是第 i 次试验的结果。

一般地，在 n 次独立重复试验中，用 X 表示事件 A 发生的次数，设每次试验中事件 A 发生的概率为 p ，则

$$P(X = k) = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n$$

此时称随机变量 X 服从二项分布(binomial distribution)，记作 $X \sim B(n, p)$ ，并称 p 为成功概率。

若 $X \sim B(n, p)$ ，则

$$\begin{aligned} E(X) &= \sum_{k=0}^n k C_n^k p^k q^{n-k} = \sum_{k=1}^n np C_{n-1}^{k-1} p^{k-1} q^{n-1-(k-1)} = np \sum_{k=0}^{n-1} C_{n-1}^k p^k q^{n-1-k} \\ &= np(p+q)^{n-1} = np \end{aligned}$$

$$D(X) = np(1-p)$$

*随机变量的均值是常数，而样本的平均值是随着样本的不同而变化的，因此样本的平均值是随机变量。

随机变量的方差是常数，而样本的方差是随着样本的不同而变化的，因此样本的方差是随机变量。

2.4 正态分布

一般地，如果对于任何实数 $a, b (a < b)$ ，随机变量 X 满足

$$\varphi_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in (-\infty, +\infty)$$

$$P(a < X \leq b) = \int_a^b \varphi_{\mu, \sigma}(x) dx$$

则称随机变量 X 服从**正态分布(normal distribution)**。正态分布完全由参数 μ 和 σ 确定，记作 $N(\mu, \sigma^2)$ 。如果随机变量 X 服从正态分布，则记为 $X \sim N(\mu, \sigma^2)$ 。

$\varphi_{\mu, \sigma}(x)$ 的图像称为**正态分布密度曲线**，简称**正态曲线**。

(参数 μ 是反映随机变量取值的平均水平的特征数，可用样本的均值去估计； σ 是衡量随机变量总体波动大小的特征数，可用样本的标准差去估计。)

标准正态分布： $X \sim N(0, 1)$

经验表明，一个随机变量如果是众多的、互不相干的、不分主次的偶然因素作用结果之和，它就服从或近似服从正态分布。

正态曲线的特点：

- (1) 曲线位于 x 轴上方，与 x 轴不相交；
- (2) 曲线是单峰的，它关于直线 $x = \mu$ 对称；
- (3) 曲线在 $x = \mu$ 处达到峰值 $\frac{1}{\sigma\sqrt{2\pi}}$ ；
- (4) 曲线与 x 轴之间的面积为 1。

* σ 越小，曲线越“高瘦”，表示总体分布越集中； σ 越大，曲线越“矮胖”，表示总体分布越分散；

若 $X \sim N(\mu, \sigma^2)$ ，则对于任何实数 $a > 0$ ，

$$P(\mu - a < X \leq \mu + a) = \int_{\mu-a}^{\mu+a} \varphi_{\mu, \sigma}(x) dx$$

该面积随着 σ 的减少而变大。这说明 σ 越小， X 落在区间 $(\mu - a, \mu + a]$ 的概率越大，即 X 集中在 μ 周围概率越大。

特别有

$$P(\mu - \sigma < X \leq \mu + \sigma) = 0.6826$$

$$P(\mu - 2\sigma < X \leq \mu + 2\sigma) = 0.9544$$

$$P(\mu - 3\sigma < X \leq \mu + 3\sigma) = 0.9974$$

在实际应用中，通常认为服从于正态分布 $N(\mu, \sigma^2)$ 的随机变量 X 只取 $(\mu - 3\sigma < X \leq \mu + 3\sigma)$ 之间的值，并简称之为**3σ原则**。

第三章 统计案例

3.1 回归分析的基本思想

回归分析(regression analysis)是对具有相关关系的两个变量进行统计分析的一种常用方法。

对于一组具有线性相关关系的数据 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$

$$\begin{cases} \hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \\ \hat{a} = \bar{y} - \hat{b} \bar{x} \end{cases}$$

其中 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ， $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ ， (\bar{x}, \bar{y}) 称为样本点的中心，回归直线过样本点的中心。

$$\text{回归方程: } \hat{y} = \hat{b}x + \hat{a}$$

线性回归模型:

$$\begin{cases} y = bx + a + e \\ E(e) = 0, D(e) = \sigma^2 \end{cases}$$

其中 a 和 b 为模型的未知参数， e 是 y 与 $bx+a$ 之间的误差。通常 e 为随机变量，称为**随机误差(random error)**。

与函数关系不同，在回归模型中， y 的值由 x 和随机因素 e 共同确定，即 x 只能解释部分 y 的变化，因此我们把 x 称为**解释变量**，把 y 称为**预报变量**。

随机误差 e 的方差 σ^2 越小，用 $bx+a$ 预报真实值 y 的精度越高。**随机误差是引起预报值 \hat{y} 与真实值 y 之间存在误差的原因之一**，其大小取决于随机误差的方差。

另一方面， \hat{b} 和 \hat{a} 为斜率和截距的估计值，它们与真实值 a 和 b 之间也存在误差，这种误差是引起预报值 \hat{y} 与真实值 y 之间存在误差的另一个原因。

由于随机误差 $e = y - (bx + a)$ ，所以 $\hat{e} = y - \hat{y}$ 是 e 的估计量。
对于样本点

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

它们的随机误差为

$$e_i = y_i - bx_i - a, i = 1, 2, \dots, n$$

其估计值为

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{b}x_i - \hat{a}, i = 1, 2, \dots, n$$

\hat{e}_i 称为相应于点 (x_i, y_i) 的 **残差(residual)**。

可以通过残差发现原始数据中的可疑数据，判断所建立模型的拟合效果。

以样本编号为横坐标，残差为纵坐标，可作出 **残差图**。

检查残差较大的样本点，确认采集该样本点过程中是否有人为错误，如有，应予以纠正，再重新利用线性回归模型拟合数据；如没有，则需寻找其它原因。

另外，对于已经获取的样本数据，

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

中的 $\sum_{i=1}^n (y_i - \bar{y})^2$ 为确定的数。因此 R^2 越大，意味着残差平方和 $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ 越小，即模型拟合效果越好； R^2 越小，残差平方和越大，即模型拟合效果越差。

R^2 表示解释变量对于预报变量变化的贡献率， R^2 越接近于 1，表示回归的效果越好。

一般地，建立回归模型的基本步骤：

- (1) 确定研究对象，明确哪个变量是解释变量，哪个变量是预报变量；
- (2) 画出解释变量和预报变量的散点图，观察它们之间的关系(如是否存在线性关系等)
- (3) 有经验确定回归方程的类型(如我们观察到数据呈线性关系，则选用线性回归方程)
- (4) 按一定规则(如最小二乘法)估计回归方程中的参数；
- (5) 得出结果后分析残差图是否有异常(如个别数据对应残差过大，残差呈现不随机的规律性等)。若存在异常，则检查数据是否有误，或模型是否合适等。

回归模型的适用范围：

- (1) 回归方程只适用于我们所研究的样本的总体；
- (2) 我们所建立的回归方程一般都有时间性；

- (3) 样本取值的范围会影响回归方程的适用范围;
- (4) 不能期望回归方程得到的预报值就是预报变量的精确值。

一般地，比较两个函数模型的拟合程度的步骤如下：

- (1) 分别建立对应于两个模型的回归方程 $\hat{y}_1 = f(x, \hat{a})$ 与 $\hat{y}_2 = g(x, \hat{b})$ ，其中 \hat{a} 和 \hat{b} 分别是参数 a 和 b 的估计值
- (2) 分别计算两个模型的 R^2 值
- (3) 若 $R_1^2 > R_2^2$ ，则模型 1 比模型 2 拟合效果更好；若 $R_1^2 < R_2^2$ ，则模型 2 比模型 1 拟合效果更好。

3.2 独立性检验的基本思想

不同的“值”表示不同类别的变量叫做**分类变量**。列出两个分类变量的频数表称为**列联表(contingency table)**。常用**等高条形图**展示列联表数据的频率特征。

利用随机变量 K^2 来判断“两个分类变量有关系”的方法称为**独立性检验(test of independence)**。

反证法原理与独立性检验原理的比较

反证法原理	在假设 H_0 下，如果推出一个矛盾，就证明了 H_0 不成立
独立性检验原理	在假设 H_0 下，如果出现一个与 H_0 相矛盾的小概率事件，就推断 H_0 不成立，且该推断犯错误的概率不超过这个小概率

一般地，假设有两个分类变量 X 和 Y ，它们的取值分别为 $\{x_1, x_2\}$ 和 $\{y_1, y_2\}$ ，其样本频数列联表(称为 2×2 列联表)为：

	y_1	y_2	总计
x_1	a	b	$a+b$
x_2	c	d	$c+d$
总计	$a+c$	$b+d$	$a+b+c+d$

假设 H_0 : X 与 Y 没有关系，即 X 与 Y 独立。

则有 $P(XY)=P(X)P(Y)$ ；

根据频率近似于概率，故有

$$\frac{a}{a+b+c+d} \approx \frac{a+b}{a+b+c+d} \times \frac{a+c}{a+b+c+d}$$

化简得 $ad \approx bc$

因此， $|ad - bc|$ 越小，两者关系越弱； $|ad - bc|$ 越大，两者关系越强；

基于以上分析，构造随机变量

$$K^2 = \frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} , \text{ 其中 } n = a + b + c + d \text{ 为样本容量}$$

K^2 的值越小则关系越小， K^2 的值越大则关系越大。(实际应用中通常要求 a, b, c, d 都不小于 5)

计算 K^2 的观测值 k 并与 K^2 作比较。

统计学研究发现，在 H_0 成立的情况下，

$$P(K^2 \geq 6.635) = 0.01$$

即在 H_0 成立的情况下， K^2 的观测值超过 6.635 的概率非常小，近似为 0.01，是一个小概率事件。

若观测值 k 大于 6.635，则有理由判定 H_0 不成立，即“ X 与 Y 有关系”。但这种判断会犯错误，犯错误的概率不会超过 0.01。

*(这里概率计算的前提是 H_0 成立，即 H_0 : 两个分类变量没有关系)

若要推断的论述为 H_1 : “ X 与 Y 有关系”。可以通过频率直观地判断两个条件概率 $P(Y=y_1|X=x_1)$ 和 $P(Y=y_1|X=x_2)$ 是否相等。如果判断它们相等，就意味着 X 和 Y 没有关系；否则就认为它们有关系。由上表可知，在 $X=x_1$ 的情况下， $Y=y_1$ 的频率为 $\frac{a}{a+b}$ ；

在 $X=x_2$ 的情况下， $Y=y_1$ 的频率为 $\frac{c}{c+d}$ 。因此，如果通过直接计算或等高条形图

发现 $\frac{a}{a+b}$ 和 $\frac{c}{c+d}$ 相差很大，就判断两个分类变量之间有关系。

利用独立性检验原理可以进一步给出推断“两个分类变量有关系”犯错误的概率。具体做法是：

(1) 根据实际问题的需要确定容许推断“两个分类变量有关系”犯错误概率的上界 α ，然后查下表确定临界值 k_0 。

$P(K^2 \geq k_0)$	0.50	0.40	0.25	0.15	0.10	0.05	0.025	0.010	0.005	0.001
k_0	0.455	0.708	1.323	2.072	2.706	3.841	5.024	6.635	7.879	10.828

(2) 利用公式计算随机变量 K^2 的观测值 k 。

(3) 如果 K^2 的观测值 k 大于判断规则的临界值 k_0 ，即 $k \geq k_0$ ，就推断“ X 与 Y 有关系”，这种推断犯错误的概率不超过 α ；否则，就认为在犯错误的概率不超过 α 的前提下不能推断“ X 与 Y 有关系”，或者在样本数据中没有发现足够证据支持结论“ X 与 Y 有关系”。

按照上述规则，把“两个分类变量之间没有关系”错误地判断为“两个分类变量之间有关系”的概率不超过 $P(K^2 \geq k_0)$ 。

定义：

$$W = \left| \frac{a}{a+b} - \frac{c}{c+d} \right|$$

则

$$K^2 = W^2 \times \frac{n(a+b)(c+d)}{(a+c)(b+d)}$$

若“X 和 Y 没有关系”则有

$$P(K^2 \geq k_0) = 0.01$$

有 $K^2 \geq k_0$ 可推出

$$W \geq \sqrt{k_0 \times \frac{(a+c)(b+d)}{n(a+b)(c+d)}}$$

即可取

$$w_0 = \sqrt{k_0 \times \frac{(a+c)(b+d)}{n(a+b)(c+d)}}$$

于是有以下判断规则：

当 W 的观测值 $w > w_0$ 时，就判断“X 和 Y 有关系”；否则，判断“X 和 Y 没有关系”。这里 w_0 为正实数，且满足在“X 和 Y 没有关系”的前提下

$$P(W^2 \geq w_0) = 0.01$$